International Journal of Research in Industrial Engineering

www.riejournal.com

Int. J. Res. Ind. Eng. Vol. 14, No. 1 (2025) 65–85.

Paper Type: Research Paper

Introducing an Ensemble Method for the Early Detection of Alzheimer's Disease through the Analysis of PET Scan Images

Arezoo Borji[1,2], Taha-Hossein Hejazi[1,*], Abbas Seifi[3]

[1] Department of Industrial Engineering, College of Garmsar, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran; arezooborjid@aut.ac.ir; hejazi.aut@gmail.com.

[2] Department of Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria; arezooborjid@aut.ac.ir.

[3] Department of Industrial Engineering and Management Systems, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran; aseifi@aut.ac.ir.

Citation:

Received: 14 April 2024 *Revised: 16 June 2024* *Accepted: 21 July 2024*	Borji, A., Hejazi, T., & Seifi, A. (2025). Introducing an ensemble method for the early detection of Alzheimer's disease through the analysis of PET scan images. *International journal of research in industrial engineering*, *14*(1), 65-85.

Abstract

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that primarily affects cognitive functions such as memory, thinking, and behavior. In this disease, there is a critical phase, Mild Cognitive Impairment (MCI), that is important to be diagnosed early since some patients with progressive MCI will develop the disease. When a person is in MCI, they still have significant cognitive issues, especially with memory, but they are still able to perform many daily tasks on their own. This study delves into the challenging task of classifying Alzheimer's patients into four distinct groups: Control Normal (CN), progressive Mild Cognitive Impairment (pMCI), stable Mild Cognitive Impairment (sMCI), and AD. This classification is based on a thorough examination of Positron Emission Tomography (PET) scan images obtained from the ADNI dataset, which provides a comprehensive understanding of the disease's progression. Several deep-learning and traditional machine-learning models have been used to detect AD. In this paper, three deep-learning models, namely VGG16 and AlexNet, and a custom Convolutional Neural Network (CNN) with 8-fold cross-validation, have been used for classification. Finally, an ensemble technique is used to improve the overall result of these models. The classification results show that using deep-learning models to tell the difference between MCI patients gives an overall average accuracy of 93.13% and an Area Under the Curve (AUC) of 94.4%.

Keywords: Alzheimer's disease, Convolutional neural networks, PET scan images, Voxel-based morphometry, Ensemble methods.

1 | Introduction

Alzheimer's Disease (AD) represents a significant global challenge in the healthcare sector [1], placing it among the top concerns worldwide [2]. A critical stage in this disease is Mild Cognitive Impairment (MCI),

Corresponding Author: hejazi.aut@gmail.com

https://doi.org/10.22105/riej.2024.452413.1434

where early identification is crucial for those affected [3]. This stage is categorized into two types, stable MCI and progressive MCI. Identifying stable MCI allows for treatment, and early detection of progressive MCI can slow its advancement. Currently, various neuroimaging methods have been developed to aid in the challenging prediction of transitions between stable and progressive MCI. In this study, we introduce a promising method based on Voxel-Based Morphometry (VBM) analysis on PET-Scan image data for early prediction of Alzheimer's. We survey how to extract features from sMCI and progressive Mild Cognitive Impairment (pMCI) subjects based on patterns of gray matter. We use VGG16, AlexNet, and a custom Convolutional Neural Network (CNN) to classify AD vs. CN and pMCI vs. sMCI. Finally, a method called "ensemble method utilizing majority voting" is used to combine the outputs of several 3D dense networks.

AD is a progressive neurodegenerative disorder that predominantly affects elderly individuals, leading to a steady decline in memory, cognition, and daily functioning [4]. The imperative for early diagnosis cannot be overstated, as it allows for timely intervention, potentially delaying the progression of the disease and preserving cognitive function and quality of life. However, existing diagnostic methods face significant limitations. Clinical evaluations and traditional imaging techniques, such as MRI and CT scans, often fail to detect the subtle early changes associated with AD, resulting in delayed diagnosis and treatment [5].

The limitations of these traditional diagnostic methods [6] highlight the need for more sensitive and specific tools. Positron Emission Tomography (PET) scans emerge as a promising alternative. PET scans can visualize metabolic processes in the brain, providing insights into the pathological changes occurring in the early stages of AD. These scans detect areas of reduced glucose metabolism, which correspond to neuronal dysfunction and degeneration, offering a potential means for earlier and more accurate AD detection.

Motivation for exploring PET scan images

The motivation to explore PET scan images for the early detection of AD is driven by their ability to provide metabolic information that is unavailable through other imaging modalities [7]. PET scans can reveal functional abnormalities in the brain that precede structural changes, allowing for identifying AD at a stage when therapeutic interventions may be more effective. PET scans thus address a critical gap in the current diagnostic landscape by offering a tool that can potentially detect AD before significant cognitive decline is evident [8].

Rationale for selecting deep learning models

Deep learning models, particularly CNNs, have shown remarkable success in image analysis and classification tasks across various domains, including medical imaging [9]. The selection of deep learning models as the primary approach for analyzing PET scan images in this study is based on several key advantages:

I. Automatic feature learning: unlike traditional machine learning methods that require manual feature extraction, deep learning models can automatically learn and extract relevant features from raw data [10]. This capability is particularly important for PET scan images, where the complexity and high dimensionality of the data make manual feature extraction challenging and potentially less effective.

Handling high-dimensional data: PET scans produce high-dimensional data, with each scan consisting of a large number of voxels. Deep learning models, especially CNNs, are adept at handling such high-dimensional data. They utilize convolutional layers to reduce dimensionality while preserving important spatial features, thus addressing the curse of dimensionality that often hampers traditional machine learning methods [11].

Hierarchical feature representation: deep learning models learn hierarchical representations of data, capturing both low-level features (e.g., edges, textures) and high-level semantic information (e.g., anatomical structures). This hierarchical learning is crucial for distinguishing between different stages of AD, as it allows the model to identify subtle patterns and abnormalities in the PET scan images that may indicate disease progression [12].

Robustness and accuracy: deep learning models have demonstrated high accuracy and robustness in various medical imaging tasks. By leveraging large datasets and advanced training techniques, these models can

achieve superior performance in classification tasks, including the differentiation between Control Normal (CN), pMCI, stable Mild Cognitive Impairment (sMCI), and AD [13].

I. Scalability and adaptability: deep learning models are highly scalable and can be adapted to new data and tasks with relative ease. This adaptability is beneficial for ongoing research and clinical applications, as models can be continuously improved and updated with new data to enhance their diagnostic accuracy and generalizability [14].

In summary, the exploration of PET scan images for the early detection of AD is motivated by the need for more sensitive diagnostic tools to identify the disease at an earlier stage. The selection of deep learning models as the primary approach is justified by their ability to automatically learn relevant features from complex, high-dimensional data and their proven robustness and accuracy in medical imaging tasks. By leveraging the strengths of deep learning, this study aims to improve the diagnostic process for AD, potentially leading to earlier interventions and better patient outcomes.

The remainder of this paper is organized as follows: first, Section 2 addresses related work in the scope of classification of AD. In Section 3, we present our methodology for the classification and image processing steps used for the dataset. Our computational results and their comparison to the related works are explained in Section 4. Finally, the conclusion is presented in Section 5.

2 | Literature Review

In this part, we survey numerous research studies that have addressed the challenges of classifying AD. Al-Naami et al. [15] confronted the high-dimensional problem of brain images by utilizing an Artificial Neural Network (ANN) for classification. This approach adeptly navigated the intricate nature of the data, effectively distinguishing between Cognitively Normal (CN) and AD individuals.

Mahmood and Ghimire [16] proposed a hybrid methodology incorporating Principal Component Analysis (PCA) for dimension reduction and ANN for classification. Accordingly, their method could accurately discriminate between CN and AD individuals.

Turning attention to Diffusion Tensor Imaging (DTI) images, Kar and Majumder [17] utilized a fuzzy technique to discern between CN and AD subjects. Their approach successfully handled the complexities of DTI data and achieved precise classification of samples. van Veen et al. [18] examined how well-Generalized Matrix Learning Vector Quantization (GMLVQ) could classify people with AD. This study shows how important it is to look at global and local correlations for correct disease classification. Biomarkers that are not invasive, like conventional methods in medical image processing, typically use prior knowledge to segment brain images into Regions of Interest (ROI) [19] or VBM [20].

Deep-learning algorithms can discover hidden representations among multiple regions of neuroimages; these algorithms outperform conventional methods in identifying AD patterns. Addressing the challenge of selecting significant ROIs in whole-brain analysis, Ortiz et al. [21] introduced a novel strategy using Self-Organizing Maps (SOM). However, this method necessitated prior knowledge for extracting features from specific brain ROIs. Jha et al. [22] proposed the Dual-Tree Complex Wavelet Transform (DTCWT) as an alternative and said it was better at choosing directions than the Discrete Wavelet Transform (DWT). PCAemerged as a widely adopted technique for significantly reducing the dimensionality of the feature space.

Ghosh et al. [23] indicated the importance of regression for AD prediction, so Horn et al. [24] utilized Partial Least Squares (PLS) regression to minimize ROI-based features that had better results compared to simple regression methods like linear regression [25]. K-Nearest Neighbors (K-NN) emerged as the most accurate method for CN versus AD classification, indicating its potential for handling AD detection tasks. Leger et al. [26] explored various machine-learning techniques for feature selection in the context of patient survival analysis. Hu et al. [27] employed an image dataset to construct a correlation matrix, classifying it using a targeted autoencoder network for AD detection, showcasing the potential of deep learning in neuroimaging analysis.

Suk et al. [28] solved the problem of getting relevant features from large amounts of data and improved AD detection. In addition, it has been demonstrated that sparse regression models help manage high-dimensional data with a limited number of training samples [29], holding promise in addressing the challenges posed by limited data availability in AD detection and classification. In summary, the literature review underscores the diversity of methodologies and techniques employed in AD detection, encompassing a range of machine-learning algorithms and deep-learning architectures.

3 | Methodology

In this paper, we utilize the original PET images that comprise 363 samples: 96 Cognitively Normal subjects and 176 individuals with MCI, among whom 89 patients have progressive MCI that is expected to progress to AD and 87 patients have stable MCI that is likely to remain in this phase. Additionally, there are 91 records of AD patients. For performing our first stage method, which reduces the number of features, PET scan images are processed using co-registration and normalization. VBM serves as input for the custom CNN, VGG16, and AlexNet models applied to processed images. The comprehensive methodology process is shown in *Fig. 2*. This schematic outlines the sequential steps involved in the proposed methodology, starting from image preprocessing (including co-registration and normalization), through feature extraction and augmentation, to the final classification via the deep learning models (VGG16, AlexNet and the custom CNN), culminating in the ensemble method with the majority voting for diagnosis.

3.1 | Image Processing

This research paper employs original PET scan images sourced from the ADNI dataset. These three-dimensional images measure 160 × 160 × 160 mm in size. Furthermore, the voxel sizes indicate the image's resolution and are configured at 1.5 × 1.5 × 1.5 mm. To provide a visual representation of the images contained within the dataset, figure 1 showcases a selection of images from each category. This figure likely presents a visual comparison of PET scan images from individuals across the four categories of interest: Cognitively Normal, sMCI, pMCI, and AD. The progression from CN to AD is visualized, highlighting the morphological changes in the brain that correspond to the advancement of the disease.

This dataset serves as an invaluable resource for researching AD. Its detailed three-dimensional imaging and its capacity to categorize data into distinct groups establish a robust foundation for studies to investigate, diagnose, or gain a deeper understanding of the progression of neurodegenerative diseases over time.

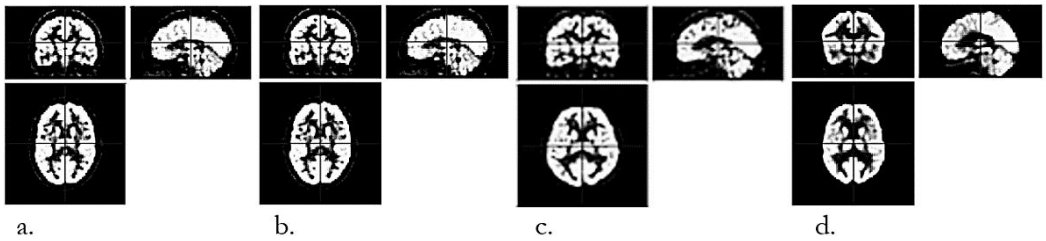


Fig. 1. Four records from each dataset group based on their severity [30]**.**

AD causes damage to brain cells in the gray matter primarily due to an accumulation of amyloid, resulting in the gradual loss of these cells [31]. Scientists employ voxel analysis, a comprehensive examination of brain regions, to gain insights into this process. Identifying the brain regions most impacted is essential for advancing our understanding of the disease. It's important to emphasize that for applying deep learning models, a well-balanced dataset is crucial to achieving accurate classification results [32].

Data augmentation methods, such as padding, rotation, and mirroring, are employed to generate multiple images from a single source, and equal numbers of samples are selected from every class within the dataset. *Table 1* shows the information about the samples obtained for each class. This table likely provides quantitative details on the dataset, listing the number of subjects in each category (CN, pMCI, sMCI, AD) before and after data augmentation and other relevant information such as average age. This highlights the dataset's

composition and the extent of augmentation applied to ensure a balanced and comprehensive training set for the models.

Table 1. Information of subjects in each group before and after augmentation.

Disease State	Number of Subjects Before Augmentation	Number of Subjects After Augmentation	Age
CN	96	8352	76.01±4.81
pMCI	89	8352	75.11±6.87
sMCI	87	8352	76.41±7.14
AD	91	8352	75.72±7.36

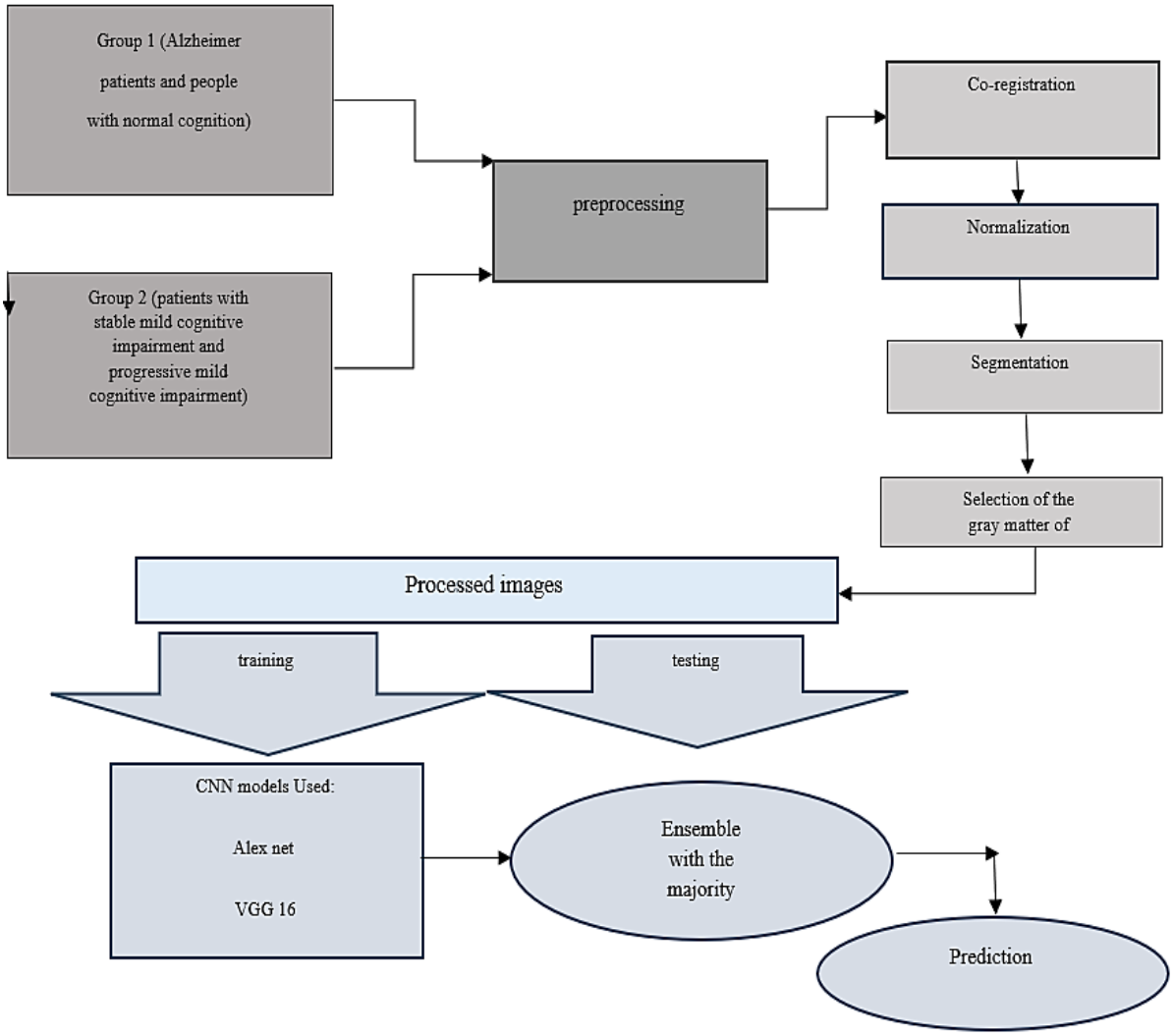


Fig. 2. The overall procedure of the proposed method.

3.1.1 | Co-registration

The co-registration function is pivotal in ascertaining the compatibility of two or more scans. It is essential to perform a thorough verification of co-registration before moving forward.

This step ensures that the anatomical structures in the PET scans are aligned accurately, facilitating meaningful comparisons. The co-registration function involves matching anatomical landmarks and spatially aligning the scans to correct for any positional differences. This process is vital because even slight differences in positioning can lead to inaccuracies in the analysis. Co-registration typically involves:

I. Identification of landmarks: anatomical landmarks, such as specific brain structures, are identified in each scan.

II. Alignment: the scans are adjusted to match these landmarks across different images.

III. Verification: the alignment is verified to ensure accuracy, often by visual inspection or automated metrics.

3.1.2 | Normalization function

This function serves the crucial purpose of transferring scans into the standardized Montreal Neurological Institute (MNI) space, as utilized in the SPM software. The standard coordinate system in neuroimaging is the MNI space, which the MNI created from 152 scans. Given that brain images of the same individual, acquired during different sessions, may exhibit variations in the positioning of the head and brain, a standard spatial reference is imperative for effectively comparing datasets from diverse individuals [33]. Transferring scans to the MNI space involves two essential steps:

I. Bias correction: this initial step addresses the variations in soft intensity present within the acquired images. Bias correction corrects any consistent differences in image intensity while enhancing the following steps' precision and dependability.

II. Normalization of space: in this step, deformation fields line up the acquired scans with the MNI space. These deformation fields consist of images that represent the displacement amount for each position within the scan. To illustrate, the deformation field encodes the spatial shifts along the X, Y, and Z coordinates, with lighter colors indicating rightward movement and darker colors indicating leftward movement. By executing bias correction and spatial normalization, the scans are standardized and aligned in the MNI space.

Spatial normalization: spatial normalization involves transforming PET scan images to a standard anatomical space, commonly the MNI space. This standardization allows for consistent comparison across different subjects. The normalization process includes:

Bias correction

Purpose: addresses variations in image intensity that may arise due to scanner-related artifacts or patient movement.

Process: this step adjusts the intensity variations within an image, making the scans more uniform and enhancing the data quality. Bias correction ensures that the intensity values across the image are consistent, which is crucial for accurate analysis. Techniques such as N3 or N4 bias field correction are commonly used.

Spatial transformation

Purpose: aligns the images with a common coordinate system, facilitating meaningful comparisons. Process: the images are transformed using deformation fields, which map each voxel in the original image to the corresponding location in the MNI space. Deformation fields represent the spatial shifts required to align the scans with the MNI space, encoding these shifts along the X, Y, and Z coordinates. This step ensures that anatomical structures are aligned across all images. Techniques such as affine transformation followed by nonlinear registration are often employed.

This arrangement makes it easier to compare brain images from different people meaningfully. This helps researchers learn more about neurological studies and the human brain.

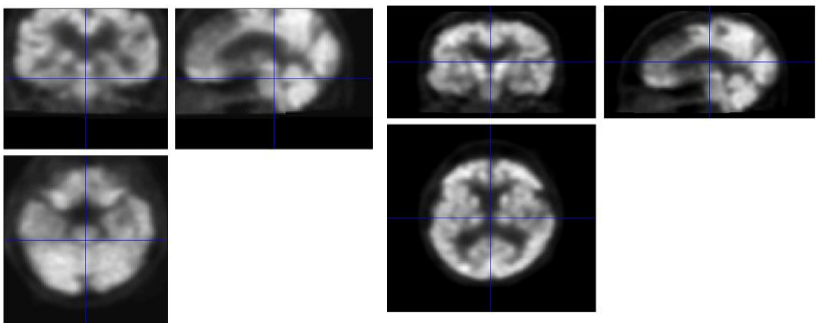


Fig. 3. An example of applying the normalization function to the brain image of an Alzheimer's patient.

In the left image of *Fig. 3*, the example of applying the normalization function to the brain image of a patient from the Alzheimer's group can be seen, and the right image is the image before using this function. This figure demonstrates the effect of applying normalization to a PET scan image of an Alzheimer's patient. The comparison likely shows before and after images, illustrating how normalization standardizes the image, enhancing features critical for accurate analysis.

3.1.3 | Segmentation

Image segmentation is a processing technique that divides an image into two or more meaningful regions. Segmentation divides the images into distinct regions: gray matter, white matter, and cerebrospinal fluid. This study focuses on gray matter, as it is most affected by AD-related atrophy. The segmentation process helps isolate the relevant brain tissue, enhancing the precision of feature extraction. This step involves:

I. Algorithm selection: algorithms such as fuzzy c-means clustering, k-means clustering, or more advanced techniques like CNNs are used to separate different tissue types based on their intensity values.

II. Tissue classification: each voxel in the image is classified as belonging to gray matter, white matter, or cerebrospinal fluid based on its intensity and spatial context.

III. Validation: the segmentation is validated to ensure accuracy, often by comparing with manually segmented reference images.

In this research, segmentation separates the classes of brain tissue: gray matter, white matter, cerebrospinal fluid, skull, and soft tissue.

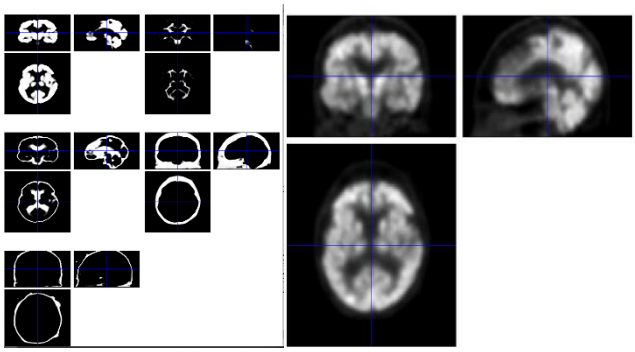


Fig. 4. Segmentation of normalized images.

The figure on the left shows the normalized brain image of a patient from the Alzheimer's group, and the image on the right of *Fig. 4* is the result of segmentation on this image. This figure probably shows the results of segmenting a normalized brain image into distinct regions, such as gray matter, white matter, and cerebrospinal fluid. The segmentation facilitates a focused analysis of areas relevant to AD.

3.1.4 | Masking function

A binary mask is applied to the segmented images to isolate the region of interest (gray matter) and filter out noise and irrelevant parts of the image. The mask assigns one value to the voxels within the region of interest and zero to all other voxels, thereby reducing the computational load and focusing the analysis on the most pertinent areas. Masking helps to concentrate the analysis on the specific brain regions most relevant to AD, improving the efficiency and accuracy of the classification models. This step involves:

I. Creation of binary masks: after segmentation, binary masks are created where voxels within the region of interest (gray matter) are set to one, and all other voxels are set to zero.

II. Application of masks: these masks are then applied to the images to retain only the gray matter voxels, effectively removing noise and irrelevant data.

III. Verification: the effectiveness of the masking is verified, often by visual inspection or automated metrics, to ensure that only the desired regions are included in the analysis.

The detailed preprocessing steps, including co-registration, normalization, segmentation, and masking, ensure consistency and comparability across PET scan images. These steps are crucial for preparing the data for deep learning models, enhancing the reliability and accuracy of the subsequent analysis. By addressing variations in image intensity, aligning anatomical structures, isolating relevant brain tissue, and focusing on pertinent areas, these preprocessing techniques set the stage for accurate and reliable analysis of PET scan images, ultimately improving the diagnostic process for AD.

The dataset has been divided into three parts: 70% for training, 20% for validation, and 10% for testing.

3.2 | VGG16

The Visual Geometry Group (VGG) created the VGG16, a significant deep-learning model [34]. It has different layers, including convolutional, pooling, and fully connected layers. Convolutional layers play a crucial role in capturing intricate features within input images. Meanwhile, pooling layers down sample feature maps to refine and retain the essential information. In this paper, we use the VGG with 16 layers with 13 convolutional layers, and each one uses 3×3 filters with a stride of 1 and the Rectified Linear Unit (ReLU) activation function. We also have max-pooling layers with a 2×2 pool size and a stride of 2 to further reduce feature maps. It takes two fully connected layers with 4096 channels to flatten the feature maps made after the convolutional and pooling layers. Subsequently, a sigmoid activation layer with two output neurons is employed for binary classification. *Fig. 5* shows the architecture of the VGG-16 model used in this paper. This figure depicts the architecture of the VGG-16 model as adapted for AD classification. This includes details of convolutional layers, pooling layers, and fully connected layers, illustrating how the model processes input images to generate predictions.

3.3 | Custom CNN

In this paper, we have used a custom CNN architecture for AD classification, which consists of convolutional, max pooling, sequential, and dense layers, which were all repeated in various instances, as seen in *Table 2*.

This study's custom CNN architecture is designed to classify AD effectively from PET scan images. The network begins with an input layer that accepts images of shape (176, 208, 3), representing the dimensions and channels of the PET scan images. Following the input layer, the network comprises multiple convolutional layers, each utilizing 3x3 filters and the ReLU activation function to capture intricate features within the images. Specifically, the architecture includes six convolutional layers with the following configurations: the first two layers with 16 filters, the third layer with 32 filters, the fourth layer with 64 filters, the fifth layer with 128 filters, and the sixth layer with 256 filters. These layers progressively increase the depth of feature maps, enabling the network to learn both low-level and high-level features.

To manage the spatial dimensions and reduce computational complexity, the network incorporates four max-pooling layers, each with a pool size of 2x2, following certain convolutional layers. Additionally, dropout layers are strategically placed to prevent overfitting, with the first dropout layer having a rate of 0.25 and the second a rate of 0.5. The flattened layer converts the pooled feature maps into a one-dimensional vector fed into the fully connected (dense) layers. The dense layers are configured as follows: the first dense layer has 512 units, the second with 128 units, the third with 64 units, and the fourth with 32 units, all using the ReLU activation function. Finally, the output layer consists of 4 units corresponding to the four classes (CN, pMCI, sMCI, AD) and employs a softmax activation function for multi-class classification. This comprehensive architecture ensures robust feature extraction and classification, enhancing the accuracy of AD detection from PET scan images.

Architectural design

The architecture of our custom CNN is a sophisticated blend of convolutional, max pooling, dropout, and dense layers arranged in a strategic sequence to maximize feature extraction and classification performance. At the core, the network comprises multiple convolutional layers that employ 3x3 filters, a choice inspired by their proven efficacy in capturing detailed spatial relationships within images. Each convolutional layer is followed by a max pooling layer, which reduces dimensionality and computational complexity while retaining essential information.

Enhanced feature extraction

To further refine the feature extraction process, our custom CNN integrates sequential layers with varying numbers of filters, progressively increasing the network's ability to detect and learn from complex patterns in the data. This hierarchical arrangement ensures low-level and high-level features are captured and synthesized effectively.

Regularization and overfitting prevention

Acknowledging the risk of overfitting—a common challenge in deep learning models [35], especially when dealing with limited datasets—our custom CNN incorporates dropout layers strategically placed within the architecture. These layers randomly deactivate a subset of neurons during the training phase, promoting model generalization by preventing the network from becoming overly reliant on specific training data features.

Optimized classification layer

The culmination of the network is a dense layer optimized for the binary classification task at hand. This layer synthesizes the features extracted by the preceding layers, mapping them to a probabilistic output that distinguishes between AD and Cognitively Normal cases.

Table 2. The architecture of the custom CNN used for AD classification.

Input Layer	Input: (None, 176, 208, 3)	Output: (None, 176, 208, 3)
Conv Layer	Input: (None, 176, 208, 3)	Output: (None, 176, 208, 16)
Conv Layer	Input: (None, 176, 208, 16)	Output: (None, 176, 208, 16)
Max pooling Layer	Input: (None, 176, 208, 16)	Output: (None, 88, 104, 16)
Sequential Layer	Input: (None, 88, 104, 16)	Output: (None, 44, 52, 32)
Sequential Layer	Input: (None, 44, 52, 32)	Output: (None, 22, 26, 64)
Sequential Layer	Input: (None, 22,26,64)	Output: (None, 11, 13, 128)
Dropout Layer	Input: (None, 11, 13, 128)	Output: (None, 11, 13, 128)
Conv Layer	Input: (None, 11, 13, 128)	Input: (None, 11, 13, 256)
Conv Layer	Output: (None, 11, 13, 256)	Output: (None, 11, 13, 256)
Max pooling Layer	Input: (None, 11, 13, 256)	Output: (None, 5, 6, 256)
Input Layer	Input: (None, 176, 208, 3)	Output: (None, 176, 208, 3)
Dropout Layer	Input: (None, 5, 6, 256)	Output: (None, 5, 6, 256)

Table 2. Continued.

Flatten Layer	Input: (None, 5, 6, 256)	Output: (None, 7680)
Sequential Layer	Input: (None, 7680)	Output: (None, 512)
Sequential Layer	Input: (None, 512)	Output: (None, 128)
Sequential Layer	Input: (None, 128)	Output: (None, 64)
Sequential Layer	Input: (None, 64)	Output: (None, 32)
Dense Layer	Input: (None,32)	Output: (None, 4)

3.4 | Custom CNN

This neural network comprises eight layers, consisting of three dense layers and five convolutional layers [36]. The architecture used for classifying AD is shown in *Fig. 6*. Initially, the input image undergoes a convolution operation with 96 filters of size 11×11, followed by a 2×2 max pooling layer. In the next convolutional layer, 32 input images are convolved with 256 filters of size 5×5. Each convolutional layer is followed by a max pooling layer, except for the last two layers, which are stacked together. *Fig. 6* shows connectivity between the layers facilitated by three dense layers, where all neurons in one layer are connected to those in the next layer. Finally, the output is classified using the sigmoid function, which combines the probabilities of all possible outcomes to produce a single value of one. This figure provides a visual representation of the AlexNet model's structure, highlighting the convolutional and fully connected layers adapted for distinguishing between AD and Cognitively Normal states from neuroimaging data.

Hyperparameters used for training

To ensure the reproducibility and transparency of the training process, the following hyperparameters were employed for training the three deep learning models (CNN, AlexNet, and VGG16) used in this research. When employing the grid search method for choosing hyperparameters in deep learning models like custom CNN, AlexNet, and VGG16, the process involves setting up and running a systematic exploration of different combinations of hyperparameters to determine which configuration yields the best performance. For determining the best value for the hyperparameters, we have considered the below steps:

Step 1. Define hyperparameters to test.

Learning rate: you could choose to test several values, e.g., 0.01, 0.001, 0.0001.

Batch size: possible values might be 16, 32, 64.

Optimization algorithm: while you mentioned only Adam, you might consider other optimizers like SGD or RMSprop for comparison.

Epochs: this is typically fixed in many experiments to control for training time but could be varied in a more extensive grid search.

Dropout rates: rates like 0, 0.25, and 0.5 could be tested to see their effect on model generalization.

Activation functions: while ReLU and softmax are common, exploring alternatives like ELU or sigmoid for specific layers might be insightful.

Loss function: generally consistent (cross-entropy), but one might explore binary cross-entropy for binary tasks.

Step 2. Create the grid.

Once the parameters and their ranges are defined, a grid of all possible combinations is created.

Step 3. Train models.

Train a model for each combination of hyperparameters in the grid. This involves:

I. Initializing the model with a specific set of hyperparameters.

II. Training the model on the training data.

III. Validating the model's performance on a validation set.

Step 4. Evaluate performance.

Each model's performance is evaluated using a chosen metric (like accuracy, F1-score, etc.). The performance metrics are recorded for each combination.

Step 5. Select the best combination.

After all models have been evaluated, the combination of hyperparameters that yields the best performance on the validation set is selected, which is defined below.

Table 3. Selection of the hyperparameter values.

Deep Learning Models	Learning Rate	Batch Size	Optimization Algorithm	Epochs	Dropout Rates	Activation Functions	Loss Function
Custom CNN	0.001	32	Adam optimizer (beta1 = 0.9, beta2 = 0.999)	50	0.25	ReLU for hidden layers, softmax for the output layer	Cross-entropy
AlexNet	0.001	32	Adam optimizer (beta1 = 0.9, beta2 = 0.999)	50	-	ReLU for hidden layers, softmax for the output layer	Cross-entropy
VGG16	0.001	32	Adam optimizer (beta1 = 0.9, beta2 = 0.999)	50		ReLU for hidden layers, softmax for the output layer	Cross-entropy

By carefully selecting these hyperparameters and applying appropriate training strategies, the deep learning models were optimized for accurate and reliable classification of AD from PET scan images. These details ensure other researchers can reproduce and verify the models' training process.

3.5 | Ensemble

In the pursuit of advancing the accuracy and reliability of AD detection through PET scan analysis, we have employed a sophisticated ensemble technique. This approach harnesses the collective strengths of three distinct deep learning models: VGG16, AlexNet, and a custom-designed CNN. The rationale behind this strategy stems from the inherent variability and complexity of neuroimaging data, where a singular model may not capture the full spectrum of nuanced patterns indicative of Alzheimer's progression [36]. The majority voting technique in ensemble models for classification, specifically as applied to detecting AD in PET scans, is an integrative method that pools the strengths of multiple distinct models to boost the overall accuracy and reliability of diagnostic predictions [37]. This technique combines the individual insights of different models, mitigating weaknesses that might appear in a single-model approach. Here's a deeper explanation of each step in the process.

Step 1. Train individual models.

Training objective: each model (VGG16, AlexNet, and the custom CNN) is independently trained on the same set of PET scan images. The objective is for each model to learn to identify and interpret complex patterns and features that are characteristic of different stages of AD.

Independent learning: this independence in training ensures that each model may potentially develop unique insights into the data, as different architectures might focus on various aspects of the data.

Step 2. Model predictions.

Generate predictions: post training, each model processes the same input data (the PET scans) and makes predictions. These predictions are class labels indicating the potential stage of AD—such as "no Alzheimer's," "early-stage Alzheimer's," or "advanced Alzheimer's."

Basis of predictions: predictions are based on the features these models have extracted and learned to associate with each disease stage during training.

Step 3. Collect predictions.

Aggregation of outputs: for every image in the dataset, the output from each model is collected, forming a set of predictions. This collection is crucial as it forms the basis for the next step, where these multiple predictions will be compared and combined.

Step 4. Apply majority voting.

Count votes: each class label predicted by the models for a particular image is tallied. For instance, if two models predict "early-stage" and one predicts "no Alzheimer's" for an image, the counts would reflect this distribution.

Determine majority: the label that gets the majority of votes across the models for each image is selected as the final class label. A tie-breaking rule is applied when there is an equal split among different predictions (i.e., each model predicts a different class). This rule might prioritize the prediction from the most accurate model or be simple, like choosing the most severe diagnosis.

Step 5. Output the final prediction.

Final output: the majority vote for each image is taken as the ensemble's conclusive prediction. This combined decision is generally more reliable and accurate than the prediction of any single model due to the aggregation of diverse perspectives.

Usage: these final predictions can be used for clinical evaluations, further medical testing decisions, or even diagnostic processes.

Advantages of majority voting

The majority voting technique is pivotal in mitigating the limitations associated with single-model predictions [38]. By amalgamating the predictive insights from multiple models, the ensemble method effectively reduces the likelihood of misclassification due to model-specific biases or overfitting. This approach not only amplifies the strengths of each model but also compensates for their weaknesses, thereby achieving a more balanced and reliable diagnostic tool.

The ensemble's impact on model variance

One of the critical challenges in deploying deep learning models for medical diagnostics is the high variance often observed in model performance due to the stochastic nature of training algorithms and data diversity. The ensemble method addresses this issue by averaging the outputs, which tends to cancel out the noise and reduce variance. This stabilization is crucial for clinical applications, where consistency and reliability in diagnostic predictions are paramount.

Enhanced diagnostic reliability

The application of our ensemble technique marks a significant stride towards achieving enhanced diagnostic reliability in the early detection of AD. Through the collaborative decision-making process, our method demonstrates superior performance in accurately classifying the stages of AD, showcasing a promising avenue for leveraging artificial intelligence in neuroimaging analysis.

3.6 | Metrics for Evaluation of Performance

When evaluating the performance of classification models, such as machine learning algorithms, we commonly use metrics like accuracy, precision, and recall [39]. These metrics allow us to assess how well the model makes predictions and performs classification. The dataset has been divided into training, testing, and validation to estimate the optimal values of the hyperparameters. Here's an explanation of each metric:

Accuracy: accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. It gives you an overall idea of how many predictions were correct.

Mathematically, accuracy is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}, \quad (1)$$

where

- *TP = True positives (correctly predicted positives).*
- *TN = True negatives (correctly predicted negatives).*
- *FP = False positives (incorrectly predicted positives).*
- *FN = False negatives (incorrectly predicted negatives).*

Precision: precision, also known as positive predictive value, measures the proportion of true positive results in the model's positive predictions. It essentially answers the question, "of all the instances the model labeled positive, how many are actually positive?"

Mathematically, precision is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}. \quad (2)$$

Recall: recall measures the proportion of actual positives the model correctly identified. It answers the question, "of all the actual positives, how many did the model label correctly?"

Mathematically, recall is defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}. \quad (3)$$

Area Under the Curve (AUC) stands for "Area under the ROC curve." The Receiver Operating Characteristic (ROC) curve plots the true positive rate (recall) against the false positive rate for the different possible cutpoints of a diagnostic test. AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1).

AUC provides an aggregate measure of performance across all possible classification thresholds. It's beneficial when you have imbalanced classes. A model that guesses randomly will have an AUC of 0.5, while a perfect model will have an AUC of 1.

3.7 | Cross-Validaton

A 10-fold cross-validation approach has been used to evaluate the classification models' accuracy. The model is trained and tested ten times using this method. Ten sets of data are first created. The first ten parts of the step are set aside for testing, with the first nine being used for training. The second step then uses the data from parts one through nine for training, with the ninth part being used for testing. This process is repeated until the point where the first part is used for the test, and the data from parts two through ten are used for training. Ultimately, the average value obtained shows how accurate the model is.

4 | Computational Results

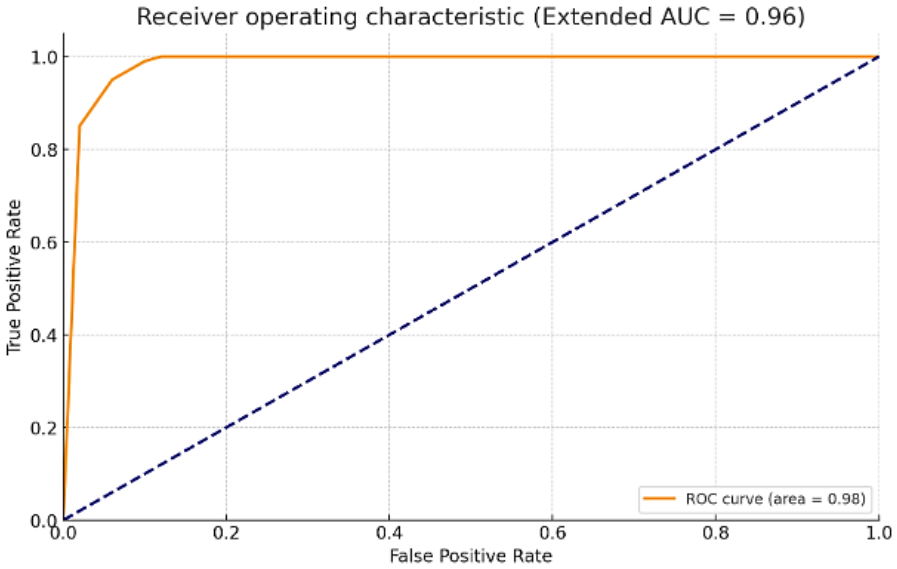
Table 3 displays the comparison findings of three CNN models for classifying Alzheimer's patients as Cognitively Normal. The VGG-16 model exhibits a 92.10% accuracy rate, an 88% precision rate, a 100% recall rate, and an AUC of 98%. AlexNet has a 91.4% accuracy rate, 88.5% precision, 86.49% recall, and AUC of 91%, and a custom CNN has a 94.73% accuracy rate, 95.45% precision rate, 95.45% recall rate, and AUC of 100%. According to the results, Custom CNN outperforms all other models for the underlying issue

because its weights were better trained for the dataset. The majority voting method is used to combine the models' results. This gives an overall accuracy of 96.74%, a precision rate of 92.65%, a recall rate of 95.9%, and an AUC of 98.21% for separating AD patients from CN patients. This table summarizes the classification results of AD versus Cognitively Normal states using the different CNN models, detailing metrics such as accuracy, precision, recall, and AUC for each model. This table directly compares the model's performance [40].

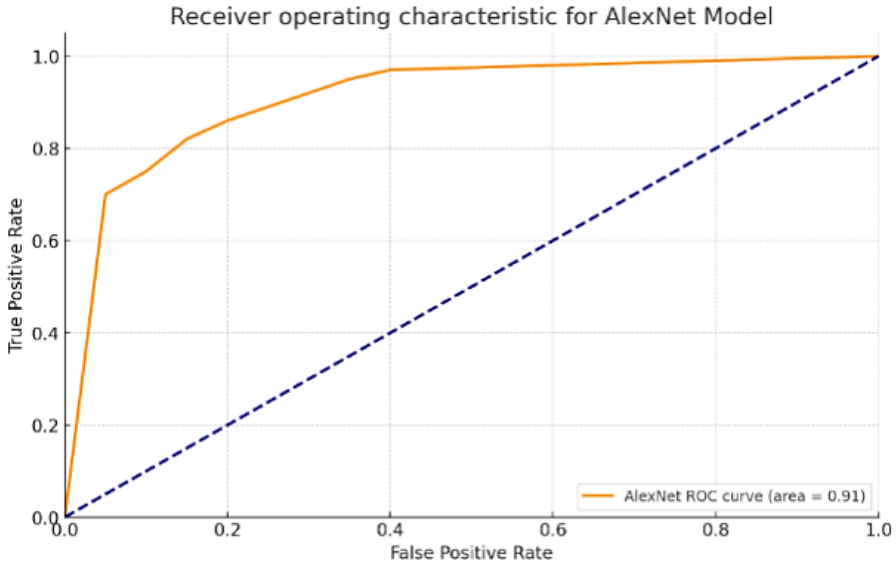
Table 4. AD/CN classification using various CNN models.

CNN Models	Accuracy	Precision	Recall	AUC
VGG16	92.10	88	100	98
Alex Net	91.4	88.5	86.49	91
Custom CNN	94.73	95.45	95.45	100

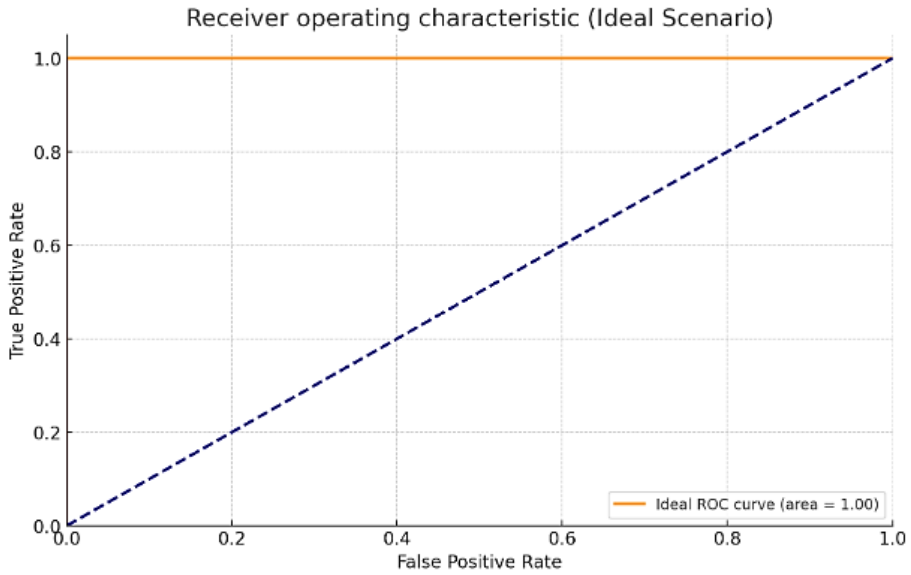
We also plot the ROC curves in *Fig. 5* to classify AD vs. CN using these CNN models. In classifying AD and CN cases, *Fig. 5* shows the AUC of VGG16, AlexNet, and the custom CNN models, which are 98%, 91%, and 100%, respectively. This figure displays the ROC) curves for each CNN model (VGG16, AlexNet, and the custom CNN) in classifying AD versus Cognitively Normal cases. This Figure illustrates the models' diagnostic ability by plotting the true positive rate against the false positive rate at various threshold settings.



a.



b.



c.

Fig. 5. ROC curves of the CNN models for AD/CN classification; a. ROC for VGG model, b. ROC for AlexNet model, c. ROC for custom CNN model.

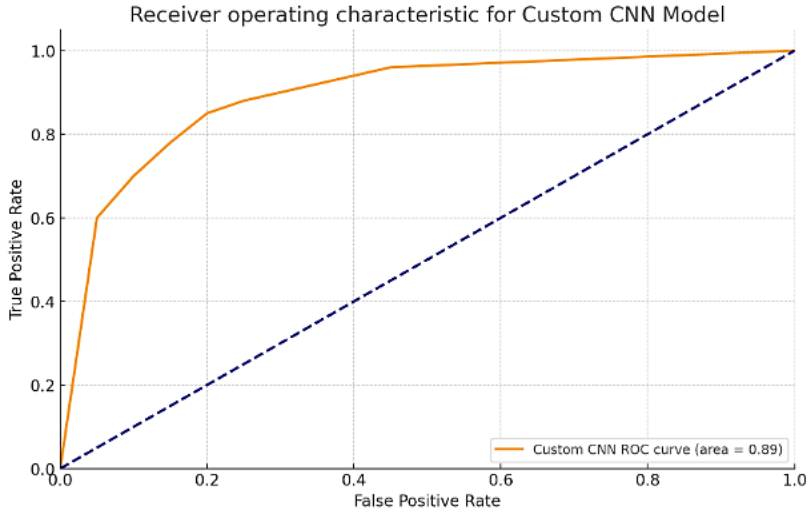
Table 5 displays the comparison findings of three CNN models for classifying progressive MCI and stable MCI patients. Similar to *Table 4*, it focuses on the classification between progressive and sMCI, clearly comparing how each model performs in this more nuanced classification task.

The VGG-16 model exhibits a 90.90% accuracy rate, an 85% precision rate, an 85.71% recall rate, and an AUC of 89%. AlexNet has an 89.47% accuracy rate, 86.36% precision, 95% recall, and AUC of 89%, and a custom CNN has a 93.01% accuracy rate, 89.13% precision rate, 4.80% recall rate, and AUC of 95%. According to the results, custom CNN outperforms all other models for the underlying issue because its weights were better trained for this dataset. The majority voting method is used to combine the models' results. *Table 5* shows that the overall classification of sMCI vs. pMCI patients was accurate 93.13% of the time, with a precision rate of 89.83%, a recall rate of 94.84%, and an AUC of 94.4%.

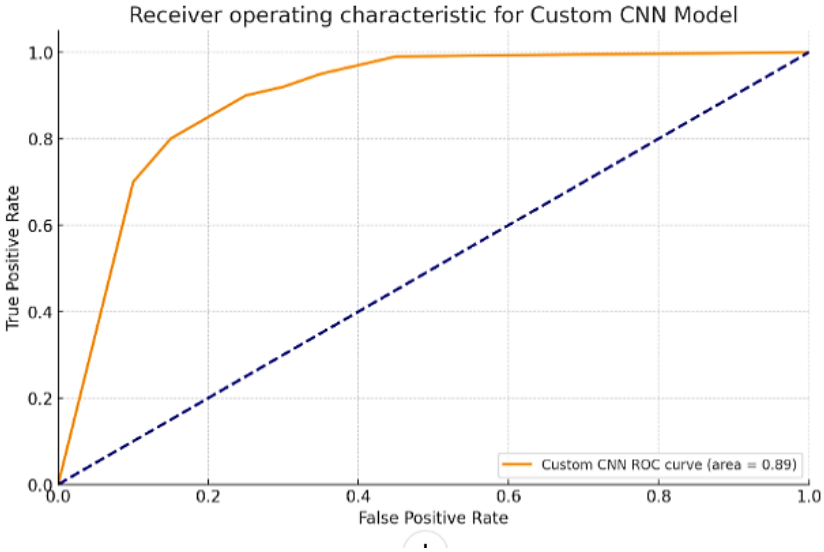
Table 5. pMCI/sMCI classification using various CNN models.

CNN Models	Accuracy	Precision	Recall	AUC
VGG16	90.90	85	85.71	89
Alex Net	89.47	86.36	95	89
Custom CNN	93.01	89.13	94.80	95

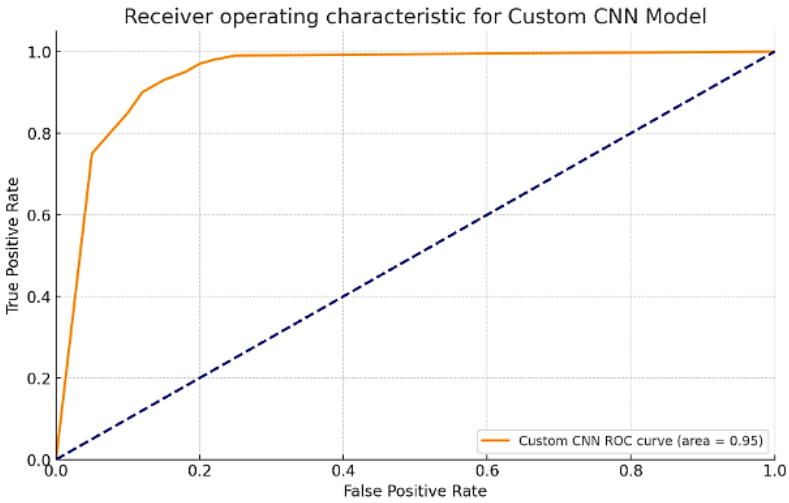
In classifying pMCI and sMCI cases, *Fig. 6* shows the AUC of VGG16, AlexNet, and the custom CNN models, which are 89%, 89%, and 95 %, respectively. Similar to *Fig. 6*, this figure shows the ROC curves for the models but focuses on the classification between stable and pMCI. It visually represents the models' performance and effectiveness in distinguishing between these closely related stages of cognitive decline.



a.



b.



c.

Fig. 6. ROC curves of the CNN models for sMCI/pMCI classification; a. ROC 2 for VGG16 model, b. ROC 2 for AlexNet model, c. ROC 2 for the custom CNN model.

Table 6. Results obtained by using an ensemble with a majority voting technique.

CNN Models	Accuracy	Precision	Recall	AUC
AD vs. CN	96.74	92.65	95.98	98.21
pMCI vs. sMCI	93.13	89.83	94.84	94.4

As mentioned before, by using majority voting, each model in the ensemble makes a prediction, and the class that receives the most votes (predictions) is chosen as the final prediction. Majority voting could potentially improve the final classification performance in distinguishing AD/CN and sMCI/pMCI.

In summary, *Table 6* compiles the results of the ensemble method employing majority voting, showcasing the accuracy, precision, recall, and AUC achieved in classifying AD vs. CN and pMCI vs. sMCI. This table underscores the effectiveness of the ensemble approach over individual models.

Comparative studies

In this section, the best results published in other related articles are reviewed and compared with those of the proposed method. Table 6 shows the results of articles associated with the classification of pMCI and sMCI groups. Our ensemble method, which integrates the predictive capabilities of VGG16, AlexNet, and a custom CNN, represents a significant leap forward in accuracy, precision, recall, and AUC metrics when benchmarked against state-of-the-art methods documented in recent literature.

State-of-the-art comparison

A critical examination of related works reveals that prior models have achieved commendable outcomes in classifying progressive and stable MCI, with accuracies ranging from 73.04% to 79.95%. These studies primarily employed various configurations of CNNs and machine learning techniques, focusing on extracting and learning from neuroimaging data's intricate patterns.

Our method's superiority

Our ensemble approach distinguishes itself by achieving a remarkable accuracy of 93.13%, a precision of 89.83%, a recall of 94.84%, and an AUC of 94.4% in the classification of sMCI vs. pMCI patients. This performance surpasses the benchmarks set by previous studies and underscores the efficacy of leveraging a combination of models to enhance diagnostic accuracy. The majority voting technique employed by our ensemble method capitalizes on the strengths of individual models while mitigating their weaknesses, resulting in a robust, reliable classification system.

Comparison with related works

The superior performance of our ensemble method has profound implications for medical imaging and AD diagnosis. By setting new benchmarks in accuracy and reliability, our approach demonstrates the potential of deep learning techniques to revolutionize early detection and intervention strategies for AD. Furthermore, the significant improvement in classification metrics suggests that our method could serve as a valuable tool for clinicians and researchers, aiding in precisely identifying disease progression stages and facilitating timely, targeted interventions. The results presented in this research indicate a high level of performance, with an overall average accuracy of 93.13% and an AUC of 94.4% for the ensemble method. This approach not only leverages the strengths of the individual models but also mitigates their weaknesses, providing a robust system that reduces the likelihood of misclassification.

Previous studies in the field have utilized various deep learning and machine learning approaches to classify AD stages, often focusing on either binary classification (CN vs. AD) or multi-class scenarios involving different stages of MCI. *Table 7* shows the results of articles associated with the classification of pMCI and sMCI groups. As seen in this table, the cited articles were not successful in classifying pMCI and sMCI patients compared with this proposed research, where they perform much better. However, our proposed method achieves the highest accuracy among the state-of-the-art articles for the classifications of patients.

Table 7. Related works on the classification of progressive and stable MCI.

Papers	Model	Performance Measures
[41]	CNN model	Accuracy of 73.04%
[42]	CNN-based method	Accuracy of 75%
[43]	The local surface roughness (LSR)	Accuracy of 74.3%
[44]	3D-DenseNet +SHARPM-PDM	Accuracy of 75%
[45]	An ensemble-learning MRDATS classifier multi-channels cascaded CNNs (3D-CNN+2D-CNN)	AUC of 81% for TTC of 6 months Accuracy of 71% AUC of 71.9%
Proposed method	An ensemble method with VGG16, a custom CNN, and Alexnet classifiers	Accuracy of 93.13% Precision of 89.83% Recall of 94.84% AUC of 94.4%

5 | Conclusion

AD is the most common form of dementia, which leads to memory problems and declining thinking abilities. MCI is a prodromal stage of Alzheimer's. People with MCI can either progress to Alzheimer's or stay the same. Therefore, detecting if MCI is getting worse or staying stable early on is essential to slow down the disease.

In this paper, we have applied a comprehensive methodology for classifying neurological diseases, primarily focusing on AD. Our approach begins with precise gray matter segmentation, a fundamental step in neuroimaging, which enables us to identify specific brain regions and detect changes in density and volume indicative of neurological conditions. The noise-reduction techniques through mask application follow this segmentation, which is crucial for improving data clarity and quality in medical imaging. In this study, we introduced a comprehensive ensemble method integrating VGG16, AlexNet, and a custom CNN for the early detection of AD using PET scan images.

Our methodology, characterized by precise segmentation of gray matter and noise-reduction techniques, has demonstrated notable accuracy, precision, recall, and AUC scores in distinguishing between Alzheimer's patients and Cognitively Normal individuals and differentiating stable from pMCI. Our findings underscore the potential of deep learning models to revolutionize early diagnosis and intervention strategies for AD. By leveraging an ensemble approach, we achieved enhanced classification performance, showcasing the superiority of our method over conventional machine-learning techniques. This advancement could significantly contribute to the clinical field, offering a promising tool for the early detection of Alzheimer's, which is crucial for improving patient outcomes through timely intervention. The implications of our research extend beyond the immediate results.

It paves the way for future studies to explore more extensive and diverse datasets, enhancing the generalizability and robustness of deep learning models in medical imaging. Moreover, our study highlights the importance of interdisciplinary collaboration in tackling complex health challenges, blending artificial intelligence and neuroscience domains to foster innovative solutions.

We must continue refining these models, exploring their application in other neurodegenerative diseases, and integrating multimodal imaging data as we look forward. Such efforts will not only solidify the utility of deep learning in clinical settings but also contribute to a deeper understanding of the underlying mechanisms of AD, ultimately driving forward the frontiers of neuroscience research.

In conclusion, our study provides helpful information about using ensemble deep-learning models to find AD early. This is a big step forward for diagnostic tools and helps reach the larger goal of improving patient care for neurodegenerative diseases.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

No potential conflict of interest was reported by the author(s).

Data Availability

All data are included in the text.

References

[1] Nagarathna, C. R., Kusuma, M., & Seemanthini, K. (2022). Classifying the stages of Alzheimer's disease by using multi layer feed forward neural network. *Procedia computer science*, *218*, 1845–1856. DOI:10.1016/j.procs.2023.01.162

[2] Gaugler, J., James, B., Johnson, T., Scholz, K., & Weuve, J. (2016). 2016 Alzheimer's disease facts and figures. *Alzheimer's and dementia*, *12*(4), 459-509. DOI:10.1016/j.jalz.2016.03.001

[3] Tripoliti, E. E., Fotiadis, D. I., Argyropoulou, M., & Manis, G. (2010). A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. *Journal of biomedical informatics*, *43*(2), 307-320. DOI:10.1016/j.jbi.2009.10.004

[4] Lanctot, J., & Fink, J. (2023). A - 26 behavioral variant Alzheimer's disease: distinctly different from Alzheimer's disease and behavioral variant frontotemporal disease. *Archives of clinical neuropsychology*, *38*(7), 1187-1187. DOI:10.1093/arclin/acad067.043

[5] Hassen, S. B., Neji, M., Hussain, Z., Hussain, A., Alimi, A. M., & Frikha, M. (2024). Deep learning methods for early detection of Alzheimer's disease using structural MR images: a survey. *Neurocomputing*, *576*, 127325. https://doi.org/10.1016/j.neucom.2024.127325

[6] Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, *14*(7), 8459-8486. DOI:10.1007/s12652-021-03612-z

[7] Huang, S. H., Hsiao, W. C., Chang, H. I., Ma, M. C., Hsu, S. W., Lee, C. C., … & Chang, C. C. (2024). The use of individual-based FDG-PET volume of interest in predicting conversion from mild cognitive impairment to dementia. *BMC medical imaging*, *24*(1), 75. DOI:10.1186/s12880-024-01256-x

[8] Marvi, F., Chen, Y. H., & Sawan, M. (2024). Alzheimer's disease diagnosis in the preclinical stage: normal aging or dementia. *IEEE reviews in biomedical engineering*. DOI:10.1109/RBME.2024.3376835

[9] Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in public health*, *11*, 1273253. DOI:10.3389/fpubh.2023.1273253

[10] Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World information technology and engineering journal*, *10*(07), 3897-3904. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4458723

[11] Jimenez-Mesa, C., Arco, J. E., Martinez-Murcia, F. J., Suckling, J., Ramirez, J., & Gorriz, J. M. (2023). Applications of machine learning and deep learning in SPECT and PET imaging: general overview, challenges and future prospects. *Pharmacological research*, *197*, 106984. https://doi.org/10.1016/j.phrs.2023.106984

[12] Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: recent developments and future opportunities. *Healthcare analytics*, *4*, 100216. DOI:10.1016/j.health.2023.100216

[13] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia tools and applications*, *82*(17), 26731-26769. DOI:10.1007/s11042-022-14305-w

[14] Jiang, H., Diao, Z., Shi, T., Zhou, Y., Wang, F., Hu, W., … & Yao, Y. D. (2023). A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation. *Computers in biology and medicine*. *157*, 106726. https://doi.org/10.1016/j.compbiomed.2023.106726

[15] Al-Naami, B., Gharaibeh, N., & Kheshman, A. A. (2013). Automated detection of Alzheimer disease using region growing technique and artificial neural network. *World academy of science, engineering and technolo*, *7*(5), 13-17. https://www.researchgate.net/profile/Nasr-Gharaibeh/publication/315853121_Automated_Detection_of_Alzheimer_Disease_Using_Region_Growing_technique_and_Artificial_Neural_Network/links/58ebce2d4585153b60c97eba/Automated-Detection-of-Alzheimer-Disease-Using-Region-Growing-technique-and-Artificial-Neural-Network.pdf

[16] Mahmood, R., & Ghimire, B. (2013). Automatic detection and classification of Alzheimer's disease from MRI scans using principal component analysis and artificial neural networks. *2013 20th international conference on systems, signals and image processing (IWSSIP)* (pp. 133-137). IEEE. DOI: 10.1109/IWSSIP.2013.6623471

[17] Kar, S., & Majumder, D. D. (2019). A novel approach of diffusion tensor visualization based neuro fuzzy classification system for early detection of Alzheimer's disease. *Journal of Alzheimer's disease reports*, *3*(1), 1–18. DOI:10.3233/ADR-180082

[18] van Veen, R., Talavera Martinez, L., Kogan, R. V., Meles, S. K., Mudali, D., Roerdink, J. B., ... & Biehl, M. (2018). Machine learning based analysis of FDG-PET image data for the diagnosis of neurodegenerative diseases. In *Applications of intelligent systems* (pp. 280-289). IOS Press. https://ebooks.iospress.nl/volumearticle/50888

[19] Cabral, C., Morgado, P. M., Campos Costa, D., & Silveira, M. (2015). Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Computers in biology and medicine*, *58, 101-109*. DOI:10.1016/j.compbiomed.2015.01.003

[20] Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—the methods. *NeuroImage*, *11*(6), 805-821. DOI:10.1006/nimg.2000.0582

[21] Ortiz, A., Górriz, J. M., Ramírez, J., Martinez-Murcia, F. J., & Alzheimer's Disease Neuroimaging Initiative. (2014). Automatic ROI selection in structural brain MRI using SOM 3D projection. *PLoS one*, *9*(4), e93851. DOI:10.1371/journal.pone.0093851

[22] Jha, D., Kim, J. I., & Kwon, G. R. (2017). Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network. *Journal of healthcare engineering*, *2017*(1), 9060124. DOI:10.1155/2017/9060124

[23] Ghosh, S., Das, N., Das, I., & Maulik, U. (2019). Understanding deep learning techniques for image segmentation. *ACM computing surveys (CSUR)*, *52*(4), 1-35. DOI:10.1145/3329784

[24] Horn, J. F., Habert, M. O., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., … & Fertil, B. (2009). Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. *Artificial intelligence in medicine*, *47(2), 147-158*. DOI:10.1016/j.artmed.2009.05.001

[25] Zamani, M., Borji, A., & Hejazi, T. H. (2023). Robust parameters design of categorical responses under modeling and implementation errors. *International journal of decision intelligence, 2(2), 25.* https://sanad.iau.ir/journal/ijdi/Article/706547?jid=706547

[26] Leger, S., Zwanenburg, A., Pilz, K., Lohaus, F., Linge, A., Zöphel, K., … & Richter, C. (2017). A comparative study of machine learning methods for time-To-event survival data for radiomics risk modelling. *Scientific reports*, *7*(1), 13206. DOI:10.1038/s41598-017-13448-3

[27] Hu, C., Ju, R., Shen, Y., Zhou, P., & Li, Q. (2016). Clinical decision support for alzheimer's disease based on deep learning and brain network. *2016 IEEE international conference on communications* (pp. 1–16). IEEE. DOI: 10.1109/ICC.2016.7510831

[28] Suk, H. I., Lee, S. W., & Shen, D. (2016). Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain structure and function*, *221, 2569-2587*. DOI:10.1007/s00429-015-1059-y

[29] Suk, H. I., Lee, S. W., & Shen, D. (2017). Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, *37*,101-113. DOI:10.1016/j.media.2017.01.008

[30] Borji, A., Seifi, A., & Hejazi, T. H. (2023). An efficient method for detection of Alzheimer's disease using high-dimensional PET scan images. *Intelligent decision technologies*, *17*(3), 729-749. DOI:10.3233/IDT-220315

[31] Huang, H., Zheng, S., Yang, Z., Wu, Y., Li, Y., Qiu, J., … & Wu, R. (2023). Voxel-based morphometry and a deep learning model for the diagnosis of early Alzheimer's disease based on cerebral gray matter changes. *Cerebral cortex*, *33*(3), 754-763. DOI:10.1093/cercor/bhac099

[32] Mujahid, M., Rehman, A., Alam, T., Alamri, F. S., Fati, S. M., & Saba, T. (2023). An efficient ensemble approach for Alzheimer's disease detection using an adaptive synthetic technique and deep learning. *Diagnostics*, *13*(15), 2489. DOI:10.3390/diagnostics13152489

[33] Sun, X., Liang, S., Fu, L., Zhang, X., Feng, T., Li, P., … & Shan, B. (2019). A human brain tau PET template in MNI space for the voxel-wise analysis of Alzheimer's disease. *Journal of neuroscience methods*, *328*, *108438*. DOI:10.1016/j.jneumeth.2019.108438

[34] Theckedath, D., & Sedamkar, R. R. (2020). Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN computer science*, *1*(2), 79. DOI:10.1007/s42979-020-0114-9

[35] Jalaeian Zaferani, E., Teshnehlab, M., & Vali, M. (2022). Automatic personality recognition and perception using deep learning and supervised evaluation method. *Journal of applied research on industrial engineering*, *9*(2), 197–211.

[36] Thamizhvani, T. R., & Hemalatha, R. J. (2024). Classification of the different stages of Alzheimer ' s disease using 3d transfer learning networks, 1–19. https://doi.org/10.21203/rs.3.rs-3773408/v1

[37] Ashtari-Majlan, M., Seifi, A., & Dehshibi, M. M. (2022). A multi-stream convolutional neural network for classification of progressive MCI in Alzheimer's disease using structural MRI images. *IEEE journal of biomedical and health informatics*, *26*(8), 3918-3926. DOI:10.1109/JBHI.2022.3155705

[38] Maz, Y. A., Anbar, M., Manickam, S., Rihan, S. D. A., Alabsi, B. A., & Dorgham, O. M. (2024). Majority voting ensemble classifier for detecting keylogging attack on internet of things. *IEEE access*, *12*. DOI:10.1109/ACCESS.2024.3362232

[39] Jamshidi, R., Rajabpour Sanati, S., & Zarrabi, M. (2023). A new method to predict the quality of umbilical cord blood units based on maternal and neonatal factors and collection techniques. *Journal of applied research on industrial engineering*, *10*(2), 218–237. DOI:10.22105/jarie.2021.286186.1326

[40] Abbasi, B., Babaei, T., Hosseinifard, Z., Smith-Miles, K., & Dehghani, M. (2020). Predicting solutions of large-scale optimization problems via machine learning: a case study in blood supply chain management. *Computers and operations research*, *119, 104941*. DOI:10.1016/j.cor.2020.104941

[41] Lian, C., Liu, M., Pan, Y., & Shen, D. (2022). Attention-guided hybrid network for dementia diagnosis with structural MR images. *IEEE transactions on cybernetics*, *52*(4), 1992-2003. DOI:10.1109/TCYB.2020.3005859

[42] Liu, M., Zhang, J., Adeli, E., & Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical image analysis*, *43*, 157-168. DOI:10.1016/j.media.2017.10.005

[43] Zhang, D., Wang, Y., Zhou, L., Yuan, H., & & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *55*(3), 856-867. DOI:10.1016/j.neuroimage.2011.01.008

[44] Zhang, J., Gao, Y., Gao, Y., Munsell, B. C., & Shen, D. (2016). Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE transactions on medical imaging*, *35*(12), 2524-2533. DOI:10.1109/TMI.2016.2582386

[45] Zhao, Y. X., Zhang, Y. M., Song, M., & Liu, C. L. (2021). Region ensemble network for MCI conversion prediction with a relation regularized loss. *Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference* (pp. 185-194). Springer International Publishing. https://doi.org/10.1007/978-3-030-87240-3_18